

Adm Policy Ment Health (2010) 37:208–220
DOI 10.1007/s10488-009-0237-8

ORIGINAL PAPER

Methodological Challenges of Characterizing Usual Care Psychotherapeutic Practice

Ann F. Garland · Michael S. Hurlburt ·
Lauren Brookman-Frazee · Robin M. Taylor ·
Erin C. Accurso

Published online: 12 September 2009

© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract There is minimal existing research providing detailed, reliable data characterizing usual community-based psychotherapy practice, and, thus, limited established methods for such research. This article identifies methodological challenges of usual care descriptive research, including, (a) general design considerations, (b) measurement, (c) data analytic, and (d) ethical challenges. Case examples drawn from studies reported in this special issue are used to illustrate the implications, strengths, and weaknesses of different methodological decisions. Central themes include achieving an acceptable balance of scientific rigor, feasibility, and generalizable practice relevance,

as well as working collaboratively with practice partners to select and implement study methods.

Keywords Research methods · Usual care · Psychotherapy

Introduction

Despite decades of research identifying empirically supported psychotherapeutic treatment models (e.g., Chambless and Hollon 1998; Chambless et al. 1996; Eyberg et al. 2008; Kazdin and Weisz 1998, 2003), minimal research exists examining actual psychotherapeutic practice in community-based settings. Psychotherapy is a private and complex interpersonal interaction, and thus, does not lend itself easily to rigorous, reliable, or feasible measurement of treatment processes or outcomes; yet measurement is essential to inform targeted improvements. Without such information, it is impossible to identify what might be working or to estimate the actual gap between evidence-based treatment practices and mainstream practice and track progress in closing that gap. Leading researchers have identified the dearth of knowledge about usual care (UC) psychotherapy practice as one of the most glaring gaps in current mental health care research (Bickman 2000; Weisz et al. 2006). To complement the growing data on specific evidence-based practices, we need data characterizing the variability in community-based practice (processes and outcomes). This research is essential to (a) provide baseline data on UC prior to intervention efforts to improve care, (b) identify specific discrepancies between evidence-based practices and UC that are potential potent targets for care improvement interventions, and (c) identify potentially effective practices delivered in UC contexts.

A. F. Garland (✉) · M. S. Hurlburt · L. Brookman-Frazee ·
R. M. Taylor · E. C. Accurso
Child and Adolescent Services Research Center at Rady
Children's Hospital, 3020 Children's Way MC 5033, San Diego,
CA 92123, USA
e-mail: agarland@ucsd.edu

M. S. Hurlburt
e-mail: Mhurlburt@casrc.org

L. Brookman-Frazee
e-mail: Lbrookman@ucsd.edu

R. M. Taylor
e-mail: Rtaylor@ucsd.edu

E. C. Accurso
e-mail: erin.accurso@gmail.com

A. F. Garland · L. Brookman-Frazee · R. M. Taylor ·
E. C. Accurso
Department of Psychiatry, University of California, San Diego,
CA, USA

M. S. Hurlburt
School of Social Work, University of Southern California, Los
Angeles, CA, USA

Given the dearth of research in this area, there are few established methods to characterize UC psychotherapeutic practice. The articles in this special issue contribute significantly to the evidence-base on UC practice and advance our knowledge of the strengths and weaknesses of different methods. The purpose of this paper is to highlight methodological challenges and decisions that must be addressed in practice-based research examining usual psychotherapeutic care. We focus on challenges that are more specific to efforts to characterize “usual care” practice, as distinct from practice delivered in the context of a more controlled intervention trial. The types of challenges described include, (a) general design considerations; (b) measurement challenges, including decisions about what to measure and how to measure it reliably; (c) data analytic challenges; and (d) ethical challenges.

To exemplify methodological challenges, we draw upon our recent studies of UC psychotherapy practice for children with disruptive behavior problems as “case examples,” (Brookman-Frazee et al. 2009b; Garland et al. under review; Hurlburt et al. 2009) highlighting strengths and weaknesses of methodological decisions and offering recommendations for future research. More detail about these studies is found in the articles, but a summary of the goals and designs is provided here. The “Practice and Research: Advancing Collaboration” (PRAC) study aims were to (a) characterize community-based outpatient care for children ages 4–13 presenting with disruptive behavior problems, (b) examine the extent to which this care was conceptually consistent with elements of evidence-based treatment for this patient population, and (c) examine variation in longitudinal outcomes and test for associations between treatment processes and outcomes. The PRAC study utilized a longitudinal observational design, and included 100 psychotherapists and 218 children/families from six publicly funded clinics in San Diego County. Multiple methods were used to characterize service use, participants’ experiences in care, treatment processes, and multiple domains of outcomes. In addition to data extracted from administrative records, a total of 3,241 psychotherapy sessions were videotaped. A sample of 1,215 sessions randomly selected for each client to represent different phases of treatment across 16 months, was coded for psychotherapeutic strategies observed to be delivered during each session. Specifics regarding the sampling and coding methods are discussed in this paper.

The Child and Adolescent Treatment Strategies study (CATS), a smaller companion study to PRAC, focused more specifically on issues of treatment continuity, coding therapy process in all sessions up through the first 15 attended by children with disruptive behavior problems and their families. As reviewed later, the measurement system was somewhat different in the CATS study, thus providing

an example of different methodological decisions. For both studies, achievement of study goals required balanced attention to external validity (e.g., representativeness of psychotherapist and patient samples), as well as internal validity (e.g., reliable and valid methods of characterizing psychotherapy processes and outcomes) and overall feasibility.

Research-Practice Partnership: The Foundation for Practice-based Research

To maximize ecological validity and feasibility, practice-based research requires strong collaboration with community partners. One of the cited reasons for a lack of existing research characterizing UC practice is providers’ concern about the potential consequences of such research (Bickman 2000). Collaborative partnership with providers is needed to overcome potential challenges. In our studies, therapist and family recruitment, as well as data collection feasibility, was greatly facilitated by a long-standing collaborative partnership with individual providers and program managers across six community-based participating clinics (described in Garland et al. 2006b). While this partnership was essential for conducting the research and strengthening the relevance and utility of the findings, such a partnership requires negotiation and compromise on the types of research design decisions outlined in this paper.

General Design Considerations

Design decisions are based on the aims of a particular study, and are balanced by feasibility constraints, ethical considerations, and research partners’ preferences and priorities. There is an established literature to guide design decisions in clinical trial research testing the impact of a particular psychosocial intervention compared to comparison conditions (Kazdin and Nock 2003; Rubin 2005; West et al. 2008). However, despite national calls for more practice-based research (National Advisory Mental Health Council 1998; Westfall et al. 2007), methods for descriptive, practice-based studies of UC have not received as much attention. Because the purpose of descriptive, practice-based research is very different from traditional intervention efficacy research, design decisions are likely to differ greatly. The aims of the PRAC study, for example, were not to assess the impact of one treatment condition compared to another, but rather to describe the variability in UC practice, to assess consistency with relatively broadly defined elements of evidence-based practice, and to identify linkages between treatment processes and outcomes. The descriptive goal required measurement of a wide array of potential treatment processes as opposed to narrowly specified

measurement of fidelity to a particular intervention model. Likewise, the goal was to characterize care as delivered under “usual” conditions, so there was no intervention in assignment of cases (i.e., no randomization to treatment or control conditions). Rather, a single cohort was studied, with emphasis on minimizing selection bias through use of participant selection and recruitment methods designed to result in the most representative samples of clinicians and patients. Clinicians were initially randomly selected for recruitment into the study and then in subsequent years of the study, all new incoming therapists were recruited to minimize selection bias. The resulting sample was very comparable to a recent national sample of therapists providing children’s mental health care, in terms of distribution by educational level (i.e., degree obtained; Glisson et al. 2008a, b). The resulting patient sample was also representative of other public sector samples on gender distribution and most common child diagnoses (e.g., Zima et al. 2005).

Data on a variety of patient outcomes were collected to allow for a description of outcome patterns and an examination of potential links between particular treatment strategies and outcome trajectories. There are many design considerations regarding assessment of treatment outcomes including selection of constructs, measurement tools, informants, timing, and handling of missing data. These design considerations are not unique to practice-based usual care research, so they are not a focus of this article. However, one of the unique challenges of practice-based descriptive research is that investigators have less control over the timing of the intervention and the associated outcome measurement. In addition, attrition may be a greater challenge when the investigative team is not involved in the delivery of care. For example, in an intervention trial, investigators often have control over treatment delivery and can delay treatment initiation until baseline outcome variable assessments are completed. Practice-based research must align assessment with an ongoing treatment system and thus there is less control over timing. One design challenge is to define a valid baseline assessment of client/family functioning from which to assess outcome change. In our studies, the logistics of the recruitment process required flexibility in some cases and we allowed up to three visits to occur prior to baseline assessment. Though not ideal, empirical tests will determine whether a later baseline assessment is associated with a different outcome trajectory (e.g., potentially less change in clinical functioning if some change already occurred after the first few visits).

Despite different aims and overall designs, clinical trial intervention research and descriptive practice-based research both grapple with methods to measure the complex phenomenon of psychotherapy practice. Well-designed treatment intervention research requires assessment of the integrity (i.e., fidelity) of treatment delivery to

support interpretive conclusions and rule out alternative explanations for results (i.e., strengthen internal validity; Perepletchikova et al. 2007). Treatment integrity includes (a) adherence to the treatment model, (b) competence in the delivery of the model, and (c) differentiation from alternative treatments or conditions (Waltz et al. 1993). There are multiple potential methods for assessing treatment integrity, but valid, reliable measurement is usually complex and costly; in fact, few psychotherapy research studies utilize adequate methods to address treatment integrity (Perepletchikova et al. 2007). The following section reviews challenges in measuring psychotherapy processes in usual care.

Measurement Challenges

Psychotherapy process measurement challenges can be summarized simplistically in the following questions: What is to be measured? How can it be measured reliably and validly?

What is to be Measured?

One significant challenge for practice research is to identify and define the treatment process elements to measure. Given that there are no established comprehensive taxonomies for psychotherapy process elements, such elements can be identified in several ways, including (a) evidence-based treatment documentation, providing a reference point for research-based interventions, (b) a wider array of clinical literature, and/or (c) reports from providers in the field about strategies they utilize. The PRAC and CATS studies utilized all three sources to identify potential treatment process elements for measurement, seeking to balance measurement of treatment strategies that are the focus of most research studies, with measurement of a broad array of treatment strategies that may be delivered in UC settings.

Range and Type of Treatment Process Elements

Adopting a focus on a more comprehensive array of treatment process elements naturally raises tensions about the number of strategies that can reasonably be measured reliably and the corresponding level of analysis. Psychotherapy practice can be assessed at many different levels of analysis ranging from a molar-level of analysis, such as classification of an entire session according to broad theoretical orientations (e.g., psychodynamic, behavioral, family systems) to more molecular-level assessments of specific therapist verbal or nonverbal behaviors (e.g., therapists’ verbatim phrases; Heaton et al. 1995). Investigators must select a level of analysis that captures meaningful treatment

process variation and is feasible. Historically, characterization at the molar theoretical level has not been particularly useful in differentiating practice patterns or outcomes (Beutler 2002; Wampold et al. 1997). It also presents problems in operationalizing definitions because some therapeutic interventions may look similar to an observer but could be characterized differently according to different psychotherapeutic theories (Goldfried 1980). For example, components of parent training interventions such as changes in disciplinary techniques, are common to both family systems and behavioral theoretical orientations, yet might be framed differently from the two perspectives, resulting in a difference in semantics but not necessarily in actual practice. Alternatively, measurement of treatment process at the molecular level of analyses (e.g., specific use of key words by therapists, such as “reinforcement” or “counter-transference”) may be more objective because there is less inference involved, but relying on utterance of specific terms may be problematic because therapists likely use many different words to convey or deliver the same clinical strategy. It would also be unwieldy if the purpose is to characterize a representative sample of practice, and it may be premature given the lack of more basic data on therapists’ practice to guide a molecular level taxonomy. Not surprisingly, limited available research indicates that assessment at the molar versus molecular levels yields different results (Heaton et al. 1995).

Most recent research characterizing psychotherapy practice assesses practice at an intermediate level of abstraction, originally defined by Goldfried (1980) as “clinical strategies” (Bearsley-Smith et al. 2008; Chorpita et al. 2007; Garland et al. under review; Garland et al. 2006a; Hogue et al. 1998; Hurlburt et al. 2009; McLeod and Weisz, under review; Weersing et al. 2002). Clinical strategies are more operationally specific than the broad theoretical orientations from which most were derived (e.g., “using positive reinforcement”), yet broader than specific verbatim utterances. The clinical strategies level of analysis has been identified as optimal for psychotherapy outcome research (Beutler and Baker 1998). Use of the clinical strategies level of analysis is likely more informative than use of the “molar” theoretical level of analysis in that many UC therapists self-identify as “eclectic,” (i.e., drawing from multiple theoretical orientations; Baumann et al. 2006; Jensen et al. 1990; Kazdin et al. 1990), rendering characterization at the theoretical level potentially problematic. There is also considerable conceptual convergence that has arisen across independent groups (listed above) with respect to labels and operational definitions of psychotherapeutic clinical strategies, lending some content validity to the constructs.

The PRAC study used an adapted version of the Therapy Process Observational Coding System for Child

Psychotherapy—Strategies Scale (TPOCS-S: McLeod and Weisz under review; McLeod 2001) to assess clinical strategies. A group of clinicians met regularly with the research team to review the original TPOCS-S for relevance to the UC context (described in Garland et al. 2006b). The final revised PRAC TPOCS-S (Garland et al. 2008a) includes 27 clinical strategies (listed in the “Appendix”). Eighteen of the 27 were retained from the original TPOCS (with minor wording changes to clarify definitions) and nine new codes were added for the PRAC study. The nine new codes reflect therapeutic techniques and content that the UC therapists reported to be common in UC, including case management and identifying client/family strengths. The 27 strategies include therapeutic techniques (e.g., role-playing, addressing client–therapist relationship, psychoeducation), as well as treatment session content (e.g., problem-solving skills, family members’ roles). Hogue and colleagues’ (Hogue et al. 2004) also differentiate measurement of treatment “techniques” and “session focus.”

The CATS study utilized a different method to identify and define clinical strategies including review of evidence-based treatment manuals, other clinical literature from multiple theoretical perspectives, and interviews with practicing UC therapists to identify treatment process elements. The CATS project developed the Child Therapy Process Rating System (Hurlburt et al. 2009), which assesses goals/strategies and methods used by therapists. Goals/strategies were identified through observation of therapist behaviors consistent with those goals/strategies. The final CTPRS consists of 39 goal/strategy target combinations summarized in Hurlburt et al. (2009). Despite their independent construction, the CTPRS and PRAC TPOCS-S have much in common, largely due to the fact that many treatment strategies and methods are common across multiple evidence-based practices, resulting in a similar level of abstraction and content in the measurement systems.

Intensity of Strategies

In addition to defining the range and type of practice elements to assess, researchers face decisions about how to characterize the intensity of clinical strategies. In intervention trial research, there is often a presumed link between the intensity with which a particular strategy is delivered and the quality of the intervention delivered, with well specified benchmarks for acceptable intensity. Intensity is also of interest in practice research, but without prescribed treatment strategies and intensity expectations, operationally defining intensity is challenging. Intensity can be assessed most objectively by measuring time spent on a strategy, or as a function of time and “thoroughness,”

defined by Hogue et al. (1996) as “extensiveness.” The CATS study placed greater emphasis on time and the PRAC study on a balance of time and thoroughness. Thoroughness reflects the depth and “follow-through” of a clinical strategy. The PRAC TPOCS-S measurement tool assessed occurrence and intensity. Observers coded whenever one of the 27 strategies was observed during a treatment session regardless of intensity (thus assessing occurrence). At the end of the session, observers assigned an intensity rating to each of the observed strategies on a seven point continuous scale (thus assessing intensity). Operational definitions and exemplar “anchors” at low, moderate and high intensity for each therapeutic strategy were provided as reference. For example, for the content strategy “problem solving skills,” a low intensity observation would be a therapist asking a child to generate alternative responses to for a reported incident (e.g., a fight on the playground). A higher intensity observation would include, for example, generating alternative responses, plus more follow-through observed within the session, with the therapist guiding the child in evaluating various alternative responses and their consequences. This approach to measuring treatment process thus yielded data on both the breadth (number of strategies observed at any intensity) and depth (intensity) of practice strategies.

Related decisions about the minimum threshold of intensity required for a clinical strategy to be coded as delivered are also complicated. For example, when a therapist says “good job” in response to a client’s description of an event, does that meet a threshold of occurrence for the strategy “delivery of positive reinforcement?” We decided to impose a fairly low threshold of intensity to characterize UC practice strategies as comprehensively as possible.

In the CATS study, the decision was made to utilize an intensity rating scale that heavily emphasized the amount of time therapists devoted to specific treatment goals/strategies and methods. This decision was made for several reasons, including to: (a) allow for direct comparison of therapeutic intensity with what would be observed in delivery of an EBP, (b) potentially to assist in achieving higher agreement among independent observers, and (c) to complement the strategy used in the PRAC study. A potential tradeoff was evident in this decision. If intensity ratings tended to be low overall, a time-based intensity rating might have low variance and possibly suffer from lower reliability. Alternatively, this approach could make comparisons with evidence-based practice reference points easier to draw and potentially might facilitate coder agreement. The slightly different approaches utilized in the PRAC and CATS studies complemented one another by placing different emphasis on the degree to which time contributed to the intensity ratings.

Measuring “Quality” of Therapeutic Strategies

Decisions regarding assessment of “intensity” in therapeutic practice are related to an even more complex issue regarding measurement of “quality.” There is a rich history of “quality of care” research wherein quality “benchmarks” are identified, often through a combination of empirical review and expert consensus, and UC practice is assessed to determine the extent to which practice meets defined quality benchmarks (Wells et al. 1996). For example, in one of the only existing studies of the nature and quality of publicly funded out-patient care for children with psychiatric disorders, Zima et al. (2005) identified quality indicators through a process of expert consensus and then reviewed charts for 813 cases across the state of California. They determined that substantial variability existed in care quality across different dimensions of care, with many cases passing criteria for quality of initial assessment, but fewer cases passing criteria in other areas, such as medication monitoring.

The type of practice-based research we report on here is related to this work, but is also somewhat distinct in that it was designed to assess the range and variability in practice, not only to assess the extent to which practice met predetermined “quality” benchmarks. However, there is conceptual overlap in that we also examined the extent to which UC practice was at least conceptually consistent with general elements of evidence-based practices (Brookman-Frazee et al. 2009b; Garland et al. under review). We use the term “conceptual consistency” because we used a relatively broad definition of elements of evidence-based practice as opposed to more strict criteria requiring fidelity to narrowly defined practice elements. For example, one of the evidence-based elements of treatment for which we assessed was role-playing for skill development. Coders recorded the occurrence of role-playing whenever it occurred, not only if it was observed in a manner entirely consistent with a role-playing exercise in a particular evidence-based treatment protocol (i.e., it was coded any time the therapist made an attempt to have the child practice a skill in vivo). Had we decided to record only those observations of practice elements that were entirely consistent with well-specified, operationalized definitions of practice elements drawn from EB treatment protocols, our resulting descriptive data on UC would be minimal (at best) because we very rarely observed delivery of any practice element that would have met fidelity standards for an EB protocol. We did observe many practice elements that were conceptually consistent with common elements of EB practice (e.g., use of positive reinforcement with children, psychoeducation directed to parents, etc.), however, the intensity was generally low and thus, not delivered in a manner totally consistent with EB

protocols (Brookman-Frazee et al. 2009b; Garland et al. under review).

Whose Behavior to Measure?

There are many additional parameters to consider in determining what to measure in order to adequately characterize psychotherapy practice. Psychotherapy processes could be characterized by assessing just the provider's behaviors, and/or the interaction of the provider and the client(s). In our studies, the primary aim was to characterize treatment delivered by providers, not the therapist–client interactive processes. There were, however, often multiple participants in treatment sessions and we decided that it was important to indicate to whom particular therapeutic strategies were directed (e.g., parent vs. child). This is particularly challenging in child/family treatment where there is often flexibility in session participation. Some sessions include the identified child patient only, others include parents and/or other family members, and many sessions are mixed. Our methodological decision to code the target of the intervention increased the complexity of the coding task but proved to be an important distinction as therapists were observed to address different content areas with parents/caregivers than with children themselves. For example, the most common content for parents was case management, whereas this was not very commonly directed to children (Zoffness et al. 2009).

How Can Psychotherapy Practice be Measured Reliably and Validly?

Psychotherapy practice can be assessed in several different ways, as demonstrated in the studies reported in this Special Issue. Direct assessment utilizes observational data via live observation, audio- or video- recording and coding. Indirect assessment may include self-report data collected from therapists and/or clients, or review of materials/records including medical charts, billing/administrative data, etc. Each of these methods has different strengths and weaknesses. Unfortunately, few studies have used multiple methods to directly examine concordance (one exception is the Hurlburt et al. study in 2009). Direct data collection and coding is more costly than indirect assessment, but direct methods are potentially more objective (Carroll and Rounsaville 2007; Lambert and Hill 1994; Perepletchikova et al. 2007). In one study of motivational interviewing interventions, therapists self-reported regular use of a variety of treatment approaches, but observers' rating of their treatment sessions revealed very little use of many of these strategies (Carroll and Rounsaville 2007). Hurlburt et al. (2009) review potential explanations for a similar lack of concordance in child and family therapy, including

observers' inability to assess therapists' intentions and formulations, subtlety of interventions that may not be recognized by observers, and therapists' lack of distinction between goals/intended interventions and actual intervention behavior in session. Given the discrepancies across assessment methods, careful consideration must be given to the meaning of information derived from different methods.

Studies Utilizing Indirect Assessment

There are a few recent studies that have attempted to characterize UC psychotherapy practice using indirect assessment methods. Zima et al. (2005) used chart record review and pre-defined quality of care benchmarks to comprehensively assess the quality of mental health care for children in publicly funded services. They determined that the chart record method was adequate for assessing broad indicators of care (e.g., medication evaluation referral made or not), but inadequate for assessing more detailed psychotherapeutic clinical strategies delivered within sessions (Zima et al. 2005). Jensen-Doss et al. (2008) developed a chart review abstraction tool to specifically assess whether certain therapy techniques were delivered to children after providers participated in a training workshop on Trauma-focused Cognitive Behavioral Therapy (Jensen-Doss et al. 2008). They achieved strong inter-rater agreement on most of the codes assessed but also acknowledged that the validity of chart review methodology is limited by the variability in detail across therapists' progress notes regarding clinical strategies utilized in sessions. Although chart review is judged to be more objective than therapist or client self-report (Jensen-Doss et al. 2008), there are still potential demand characteristics that likely influence chart recording, somewhat reducing the objectivity. Furthermore, it can be difficult to extract information about intensity of therapeutic strategies from chart review relative to other methods.

Therapist self-report has been the most common method for assessing practice patterns and therapist attitudes/preferences in practice (Aarons 2004; Addis and Krasnow 2000; Baumann et al. 2006; Kazdin et al. 1990; Nelson et al. 2006; Sheehan et al. 2007; Weersing et al. 2002). The Therapy Process Checklist (TPC; Weersing et al. 2002) and its adapted version that includes family interventions (Baumann et al. 2006) relied on rigorous psychometric testing and development. This tool yields data on the extent to which therapists endorse 50 treatment strategies consistent with major theoretical orientations (e.g., psychodynamic, behavioral, cognitive). The tool can be used to assess general endorsement of practice patterns, or the therapeutic strategies used with a particular case. However, the extent to which it has been cross-validated to examine concordance with observers' ratings, and/or clients' experience in sessions is limited.

Bearsley-Smith et al. (2008) in Australia recently reported on the development of another therapist self-report tool called the Treatment Recording Sheet (TRS). This tool was developed in collaboration with therapists through an iterative process using qualitative methods and thus has strong ecological validity. It is designed to characterize UC treatment with adolescents in community-based care and was adapted from tools used in the State of Hawaii. The TRS includes 44 specific intervention strategies grouped into 12 categories, plus 6 related activities/actions (e.g., school liaison). Therapists are asked to designate the target of the intervention (i.e., child, parent, etc.). Therapists reportedly found the tool useful in providing a language for describing interventions. Given the lack of any established taxonomy for labeling therapeutic intervention strategies, we agree that this is an essential function provided by these types of tools (Brookman-Frazee et al. 2009a). However, the authors acknowledge that the lack of psychometric data on the measure, including tests of concordance between self-reported and observed practice, is a significant limitation in need of further research.

Studies Utilizing Direct Assessment

Fewer studies have utilized direct assessment of child/family psychotherapy in UC settings, due likely to the cost and complexity in collecting and analyzing these data. The PRAC and associated CATS studies are the first to attempt to provide a direct assessment of in-session therapist behavior for a relatively large sample of representative therapists and children/families. As described earlier, the PRAC study utilized the PRAC TPOCS-S, (Garland et al. 2008a) which includes 27 clinical strategies targeting children and/or their parents (list included in ‘Appendix’). Interestingly, considerable overlap exists between the strategies included in the PRAC TPOCS-S, the CATS CTPRS, and the strategies listed in the TRS (Bearsley-Smith et al. 2008) described above, even though they were developed independently (and in different countries); techniques such as “psychoeducation,” “goal setting,” “interpretation,” and “modeling” are common. All tools were developed in collaboration with UC therapists and reflect the heterogeneity of UC practice. Use of qualitative, collaborative methods to assemble assessment tools that are relevant and valid for UC practice has been strongly recommended (Baumann et al. 2006).

Observational ratings may be the most reliable and valid method to assess therapist behavior, but they cannot assess therapists’ cognitions or formulations of therapeutic interventions. Observation of in-session treatment is also restricted to therapeutic interventions delivered in the time and space confines of the scheduled treatment visit and thus

does not capture any interventions outside of structured sessions (e.g., telephone calls, meetings with other professionals at schools or other agencies, or even waiting-room interactions). In the PRAC and CATS studies, we elected to focus on the major intervention elements that would appear in the context of the office visits and presumed that discussions captured in the office would often reflect outside activities (e.g., case management activities; Zoffness et al. 2009). This was an appropriate and feasible decision for our studies of traditional, office-based outpatient care. However, it would not have been appropriate for other interventions that are more flexible with respect to intervention locations and scheduling (e.g., home-based or school-based interventions; Kataoka et al. 2006; Schley et al. 2008; Schoenwald et al. 2008). Observational assessment of practice delivered outside the structure of traditional office-based practice would require a somewhat different methodology.

Reliable Measurement of Practice

The ultimate validity of a direct assessment of practice depends on reliability of the coding method. In this section we review challenges to achieving reliable coding on observational practice data. Achieving strong reliability while also capturing the heterogeneity of UC practice may be more challenging than more targeted coding of the adherence to well-specified treatment techniques in clinical trial research. The reliability of coding UC practice depends on well-developed operational definitions of practice elements and adequate selection, training, and monitoring of coders. One of the challenges in identifying codeable practice elements is the “width” of the definitional boundaries. Codes with wider definitions that include a broad range of therapist behavior (e.g., use of positive reinforcement with children) tend to accumulate higher occurrence ratings, but reliability may be challenging if the definition is too diffuse. Codes with narrower definitions provide more specificity about what was actually delivered, but may result in lower occurrence and thus may also have low reliability associated with infrequent observation. We found that more rarely observed practice elements (e.g., addressing the client–therapist relationship) were often associated with lower reliability. Utilizing codes with narrower definitions also requires more total codes to describe the array of UC practice; more total codes likely reduces feasibility and limits reliability. However, restricting the total number of codes to include only high frequency elements limits the ultimate value of the results. For example, assignment/review of homework was a relatively infrequently observed element in the PRAC study, but the low frequency of this element has significant implications for EB practice.

Selection and Training of Coders

Another important decision relates to the training level and experience of psychotherapy coders. There are potential strengths and limitations in selecting experienced psychotherapists as coders. Experienced psychotherapists have a well-developed vocabulary and frame of reference for characterizing psychotherapy approaches and are familiar with distinctions across theoretical orientations, etc. However, they may be less objective in recording other therapists' interventions as viewed through their own schema of psychotherapy practice. In the PRAC study, we opted to exclude experienced psychotherapists to avoid this potential source of bias (coders were research assistants and graduate students with some limited academic background in psychology). In a large scale study requiring so much coding, it was also more cost-efficient to utilize non-psychotherapists as coders and we were able to achieve adequate reliability on most of the codes using non-therapist coders.

Additional challenges in achieving inter-rater reliability include the extent to which specific observed behaviors can be coded with multiple codes (and/or multiple targets). In reality, therapists often integrate different types of therapeutic strategies. For example, a therapist might ask a client how she felt about a particular event and then quickly move to teaching her an affect management skill such as deep breathing, demonstrating the skill, and asking the client to practice the skill. In our PRAC TPOCS-S coding system this would be coded using three therapist technique codes (psychoeducation, modeling, role-playing/practice), and two therapeutic content codes (affect education and affect management). Attending to multiple individual elements of practice that may be interwoven in a single interchange is challenging, but does represent the reality of practice (Hogue et al. 2004).

Operational definitions of therapeutic strategies for children are also complicated by variability across developmental stages. Specifically, the same type of therapeutic strategy (e.g., affect education or problem solving skills training) may be used very differently with pre-school age children compared to adolescents (e.g., the therapist may communicate the concepts more simplistically). Likewise, variability associated with different diagnostic profiles must be acknowledged. This was particularly striking in our sample of children ages 4–13 years presenting with disruptive behavior problems. To reflect the UC patient population, there were few exclusionary criteria for participants, and thus a great deal of diagnostic variation and comorbidity. Some participants had comorbid Autism Spectrum Disorders (ASD), others had mood or anxiety disorders, and many were diagnosed with Attention Deficit disorders. Delivery of a particular therapeutic strategy (e.g., problem solving skills) could look very different when

targeting a child with an ASD compared to a mood disorder. We instructed coders to *generally* take into account the child's developmental level (age, obvious developmental delay) and clinical characteristics (obvious attentional/regulation problems) when coding so that variability in therapists' delivery methods for the same therapeutic strategy could be coded accurately. For all the reasons noted above, coder training was challenging and reinforcement of coding decisions and protocols was required consistently across the project (i.e., booster sessions).

Data Analytic Challenges

Data analytic challenges include decisions regarding (a) criteria for acceptable reliability; (b) aggregating data into subscales based on empirical or theoretical criteria; (c) aggregating data across sessions, clients, and/or therapists; (d) implications of nested, multi-level data; and (e) impact of therapist turnover.

Criteria for Reliability

Given all the complexity in achieving inter-rater reliability on psychotherapy practice characteristics, it may be unrealistic to assume that reliability will be uniformly strong in this type of research. However, what should the criteria be for acceptable reliability for different analytic purposes? Standardized criteria for acceptable reliability have been published (Cicchetti 1994; Landis and Koch 1977), but reliability estimates can be calculated in different ways (e.g., aggregated across sessions vs. by item). For example, in our analyses of inter-rater reliability on the intensity scale for the PRAC study, the ICC aggregated across all codes at the session level was 0.78. However, as expected, individual item ICC's were more variable (range of 0.21–0.91, with a mean of 0.61). This illustrates that the interpretation of reliability differs based on the level of aggregation and it is more difficult to achieve high reliability on individual items. Kappa statistics were used to assess reliability for any occurrence of a treatment strategy (as opposed to scaled intensity). A similar pattern emerged, whereby the Kappa aggregated across codes at the session level was 0.67, but it ranged by code from 0.25 to 0.89, with a mean of 0.51 for individual codes, also reflecting moderate inter-rater reliability. The two codes with the lowest occurrence (observed in fewer than 15% of sessions) had the lowest reliability (kappas < 0.45). Results from the CATS study parallel those of PRAC with regard to reliability of the Child Therapy Process Rating System, including overall reliability, variability in reliability of individual treatment process codes, and lower reliabilities for infrequently occurring strategies (Hurlburt et al. 2009).

Whether utilizing ICCs or Kappa statistics, our work illustrates that it is possible to measure many psychotherapy practice elements with adequate inter-rater reliability (Cicchetti 1994; Landis and Koch 1977). Most codes have reasonably strong reliability, but variability in reliability does exist, particularly related to the frequency of element occurrence. This suggests that although it is difficult to reliably capture a wide range of therapist intervention strategies, it is possible to reliably code a relatively comprehensive array of UC therapist behavior.

Data Aggregation Decisions

The extent of data aggregation is a theme across many analytic decisions. The clearest way to present data is at the individual item level (individual clinical strategy observed); however there are many important aims/research questions that rely on aggregation of items into conceptually derived, or empirically derived subscales. The paper by Brookman-Frazee et al. (2009b) is an example. The research goal was to determine the extent to which UC therapists were delivering care consistent with previously identified common elements of evidence-based practice (EBP) for children with disruptive behavior problems and their parents (Garland et al. 2008b). In analyses in which we calculate a composite of multiple strategy items (e.g., a summary composite of mean intensity on all codes classified as “EBP Strategies”), we were more conservative in our criteria for including a specific code than in a broad description of the heterogeneity of care (Garland et al. under review). For example, the analyses reported in this issue (Brookman-Frazee et al. 2009b) included only codes that achieved a Kappa ≥ 0.40 and an ICC ≥ 0.5 and occurred in more than 1% of sessions in the EBP composite. There are many different ways that items could be grouped based on conceptual, theoretical, or empirical criteria, but the implications for reliable measurement when grouping individual items of varied reliability must be considered.

In addition to aggregation across items, decisions must be made about aggregation across sessions. There is minimal research to inform these types of decisions; little is known about consistency in practice strategies across time. There may be important patterns in the sequencing of treatment strategies that could be lost by aggregating across all sessions. The PRAC and CATS studies are designed to empirically address this issue. The CATS study examines every session for the first 15 sessions, and the PRAC study examines a random sample across 16 months of treatment.

In addition, aggregation across items and use of the mean intensity on clinical strategies may obscure important effects that could be found by selecting the highest intensity across sessions, or counting the number of strategies

observed above a specified threshold across sessions. These types of questions highlight important areas for future research, some of which can be explored in the studies presented in this Special Issue.

Implications of Nested Multi-level Data

Observational data collected from descriptive, practice-based research are likely to reflect a hierarchical structure in which observations are obtained at multiple levels and are nested within levels. For example, in our studies, observational data from individual sessions are nested within children, children are nested within therapists, and therapists are nested within clinics and organizations. There is variability in the ratios of observations at each level (e.g., number of coded sessions per child and number of children per therapist). In analyses in which therapeutic strategies at the session level are used as the dependent variable (e.g., Brookman-Frazee et al. 2009b), ICCs are calculated to estimate whether significant variance in the dependent variable is accounted for at each level of the data structure. We use the conventional ICC cutoff of 0.05 (Hox 2002) to determine whether to account for each level in subsequent analyses. These types of multi-level data are complex and rich, but they require sophisticated analytic approaches. One of the challenges is to be clear about the consistency in the levels of interest for different analytic purposes and/or different research questions. For example, in analyses in which the dependent variable is at the child level (e.g., child symptom or functioning outcomes), the session level data cannot be used as an independent variable because the independent variable cannot be at a lower level than the dependent variable. Therefore, the session-level data on treatment processes must be aggregated to the child level (see above regarding implications of cross-session aggregation). An added complexity is the variability in the number of sessions on which this aggregated data is based, but this can be accounted for by using the number of sessions as a covariate in analyses.

Impact of Therapist Turnover

Therapist turnover in UC settings is typically high (Aarons and Sawitzky 2006; Glisson et al. 2006, 2008a, b; Knudsen et al. 2003). This reality of the practice contexts contributes to data analytic challenges for multi-level analyses that include therapist characteristics, as multiple therapists per child potentially introduces an additional level to the data structure. This problem is more complex in practice-based research assessing “naturalistic” treatment processes and would be less likely to occur in a controlled treatment trial. Deleting data from subjects with more than one therapist would exclude a significant proportion of clients in UC. In

our broad descriptive analyses characterizing care, we include care provided by all therapists; however, in analyses of therapist effects associated with outcomes we had to select only one therapist per child client and selected the therapist who had the most visits with the client. This allowed us to examine how variability in therapist characteristics is associated with variability in treatment delivery or child outcomes.

Ethical Challenges

Given the highly sensitive and private nature of psychotherapy, concerns regarding informed consent to participate in observational research and safeguards for maintaining confidentiality of psychotherapy data are particularly salient. Potential risks and benefits for participating in descriptive, practice-based research are different from intervention trial research where there is the potential for receiving improved treatment (as well as risk of unanticipated harm). As detailed below, informed consent documents need to be explicit about how data will be collected, analyzed, and stored, as well as who will have access to it. Participants need to be assured that they can stop data collection whenever desired. Access to the data must obviously be well protected and those with access must be trained carefully on confidentiality procedures (e.g., prohibiting any discussion of observed practice in any setting or situation). This can become even more critical with potential dual roles, such as in the PRAC study when a research assistant coder later became a trainee therapist in one of the study sites.

In addition to over-arching concern regarding maintaining their clients' privacy, providers have also expressed anxiety about how observational data may be used to evaluate their professional performance. These issues need to be addressed explicitly in the planning phases of the study and in the informed consent documents outlining how the data will be used and to whom it will be released and reported. Our protocols established that program administrators would not have access to individual provider data on practice and that such data would not be available for any performance evaluation purposes. This seemed appropriate in an exploratory, descriptive study of psychotherapy that did not attempt to assess the "quality" of care. However, there was still the possibility of observing care that was unethical or inappropriate, and thus there was a need to establish an operational definition and threshold for when (and how) the research team would intervene if unethical or inappropriate care was observed (Garland et al. 2008). Professional ethical guidelines, as well as mandatory reporting laws support such intervention if blatant examples of therapists' abusive or grossly

negligent treatment is observed (American Psychological Association 2002). However, there is still subjective judgment regarding negligent or abusive treatment and there are methodological limitations. Specifically, for example, if a child is observed in a session describing clearly abusive parental behavior in the home, the therapist may or may not be observed following-up regarding a mandatory call to Child Protective Services. This report may likely take place outside the observed session(s), so the research team will not necessarily know if the report has been filed. Our decision in these cases where there was no observation of explicit follow-up was to check with the therapist to assure that a report had been filed. This did not result in any significant conflict or withdrawal from the study. Informed consent documents (for providers and clients) explicated that the research team would communicate with the therapist (and professional oversight boards if necessary) if any potential harm to clients was observed. Of course, such communications had to be handled sensitively to maintain a collaborative partnership because researchers did not want to be perceived as "checking up" on therapists in a critical way.

Any observational research on practice carries an additional ethical challenge related to the potential impact of the observation itself. The "Hawthorne effect" (Mayo 1933) has been well established in psychological and organizational research; observation of a phenomenon is, in itself, an intervention and can impact the phenomenon being observed (Mufson et al. 2004; Vinnars et al. 2005). There are few established methodological guidelines for minimizing this effect, but common sense suggests that use of unobtrusive measurement procedures which become routinized may help, in addition to minimizing potential consequences of data collection (e.g., performance evaluation discussion above). We utilized very small unobtrusive video cameras mounted high in providers' offices and high quality microphones on desks to minimize explicit attention to the video-taping. Support staff in the clinic settings facilitated video-tape recording of every session with consented participants, even though only a random sample of sessions was observed and coded. The goal was to minimize the data collection burden on therapists and to routinize the videotaping. There is no way to know how our observational data collection methods may have impacted practice itself, but virtually all provider participants indicated to us that the procedures did become routine. However, two providers withdrew from the study because they felt that their self-consciousness about the recording was inhibiting flexibility and spontaneity in practice. Children were rarely observed "playing to" the camera (e.g., waving or making faces), suggesting that the recording process faded into the background.

In sum, practice-based research presents some unique ethical challenges that impact research design and stakeholder partnership. These types of ethical issues need to be addressed early in the proposed study planning and revisited as specific issues arise.

Summary/Conclusions/Recommendations for Future Research

Establishing the optimal methods for practice-based research characterizing treatment requires attention to many different types of challenges ranging from ethical consideration to community partners' priorities, design and measurement challenges, and data analytic decision-making, let alone the practical constraints imposed by budgets, etc. Some of these challenges are interdependent, but many may be conflicting.

There are many resources to consult to inform methodological decisions in traditional intervention trial research, but there are fewer for practice-based descriptive research. This article is intended as a preliminary resource on which to build future research. We hope that our experience may prove useful in advancing the field.

Acknowledgments This work was supported by NIH R01MH66070 (A. G.), K23MH077584 (L. B. F.), K01MH064079 (M. H.), F31MH083399 (E. A.), and P30MH074678. The authors thank Scott Roesch, Ph.D. for methodological and statistical consultation and Rachel Haine, Ph.D. for editorial suggestions.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

Table 1 PRAC TPOCS-S codes: techniques and content categories

Techniques	Content
T1. Using positive reinforcement/rewards	C1. Cognitive restructuring
T2. Using limit-setting/punishment	C2. Problem solving/social skills
T3. Modeling	C3. Affect education
T4. Role-playing/practice	C4. Affect management
T5. Assigning or reviewing homework	C5. Principles of positive reinforcement
T6. Psychoeducation	C6. Principles of effective limit-setting/extinction/punishment
T7. Establishing/reviewing treatment goals or agenda	C7. Family members' roles

Table 1 continued

Techniques	Content
T8. Addressing client–therapist relationship	C8. Parent–child relationship
T9. Addressing client resistance	C9. Improved communication
T10. Interpreting meaning of behavior	C10. Anticipating set-backs or relapses
T11. Exploring client's/family's past	C11. Addressing child's external care/activities (case management)
T12. Using genograms	C12. Addressing parent/family issues, care
T13. Assessing problems/events	
T14. Identifying/addressing client's strengths	
T15. Play/art	

References

- Aarons, G. A. (2004). Mental health provider attitudes toward adoption of evidence-based practice: The Evidence-Based Practice Attitude Scale (EBPAS). *Mental Health Services Research*, 6, 61–74.
- Aarons, G. A., & Sawitzky, A. (2006). Organizational climate mediates the effect of culture on work attitudes and turnover in mental health services. *Administration and Policy in Mental Health and Mental Health Services Research*, 33, 289–301.
- Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology*, 68, 331–339.
- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060–1073.
- Baumann, B. L., Kolko, D. J., Collins, K., & Herschell, A. D. (2006). Understanding practitioners' characteristics and perspectives prior to the dissemination of an evidence-based intervention. *Child Abuse and Neglect*, 30, 771–787.
- Bearsley-Smith, C., Sellick, K., Chesters, J., Francis, K., & Gippsland Adolescent Depression Research Group. (2008). Treatment content in child and adolescent mental health services: Development of the treatment recording sheet. *Administration and Policy Mental Health*, 35, 423–435.
- Beutler, L. E. (2002). The dodo bird is extinct. *Clinical Psychology: Science and Practice*, 9, 30–34.
- Beutler, L. E., & Baker, M. (1998). *The movement toward empirical validation: At what level should we analyze, and who are the consumers?* (pp. 43–65). Thousand Oaks, CA: Sage Press.
- Bickman, L. (2000). The most dangerous and difficult question in mental health services research. *Mental Health Services Research*, 2, 71–72.
- Brookman-Frazee, L., Garland, A. F., Taylor, R., & Zoffness, R. (2009a). Therapists' attitudes towards psychotherapeutic strategies in community-based psychotherapy with children with disruptive behavior problems. *Administration and Policy in Mental Health and Mental Health Services*, 36, 1–12.
- Brookman-Frazee, L., Haine, R. A., Baker-Ericzen, M., Zoffness, R., & Garland, A. F. (2009b). Factors associated with use of evidence-based practice strategies in usual care youth psychotherapy (this issue).

- Carroll, K. M., & Rounsaville, B. J. (2007). A vision of the next generation of behavioral therapies research in the addictions. *Addiction*, 102, 850–862.
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7–18.
- Chambless, D. L., Sanderson, W. C., Shoham, V., Bennett Johnson, S., Pope, K. S., Crits-Christoph, P., et al. (1996). An update on empirically validated therapies. *Clinical Psychologist*, 49, 5–18.
- Chorpita, B. F., Becker, K. D., & Daleiden, E. L. (2007). Understanding the common elements of evidence-based practice: misconceptions and clinical examples. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46, 647–652.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rule of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290.
- Eyberg, S. M., Nelson, M. M., & Boggs, S. R. (2008). Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. *Journal of Clinical Child and Adolescent Psychology*, 37(1), 215–237.
- Garland, A. F., Brookman-Frazee, L., Hurlburt, M., Accurso, E. C., Zoffness, R., Haine, R. A., & Ganger, W. A. (under review). Characterizing community-based psychotherapy for children with disruptive behavior problems.
- Garland, A. F., Brookman-Frazee, L., McLeod, B. (2008a). Scoring manual for the PRAC study therapy process observational coding system for child psychotherapy: strategies scale.
- Garland, A. F., Hawley, K. M., Brookman-Frazee, L., & Hurlburt, M. S. (2008b). Identifying common elements of evidence-based psychosocial treatments for children's disruptive behavior problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47, 505–514.
- Garland, A. F., Hurlburt, M. S., & Hawley, K. M. (2006a). Examining psychotherapy processes in a services research context. *Clinical Psychology: Science and Practice*, 13, 30–46.
- Garland, A. F., McCabe, K. M., & Yeh, M. (2008c). Ethical challenges in practice-based mental health services research: Examples from research with children and families. *Clinical Psychology: Science and Practice*, 15, 118–124.
- Garland, A. F., Plemmons, D., & Koontz, L. (2006b). Research-practice partnerships in mental health: Lessons from participants. *Administration and Policy in Mental Health and Mental Health Services*, 33, 517–528.
- Glisson, C., Dukes, D., & Green, P. (2006). The effects of the ARC organizational intervention on caseworker turnover, climate, and culture in children's service systems. *Child Abuse and Neglect*, 30, 855–880.
- Glisson, C., Landsverk, J., Schoenwald, S., et al. (2008a). Assessing the organizational social context (OSC) of mental health services: Implications for research and practice. *Administration and Policy in Mental Health and Mental Health Services Research*, 35, 98–113.
- Glisson, C., Schoenwald, S. K., Kelleher, K., Landsverk, J., Hoagwood, K. E., Mayberg, S., & Green, P. (2008b). Therapist turnover and new program sustainability in mental health clinics as a function of organizational culture, climate, and service structure. *Administration and Policy in Mental Health and Mental Health Services Research. Special Issue: Improving Mental Health Services*, 35, 124–133.
- Goldfried, M. R. (1980). Toward the delineation of therapeutic change principles. *American Psychologist*, 35, 991–999.
- Heaton, K. J., Hill, C. E., & Edwards, L. A. (1995). Comparing molecular and molar methods of judging therapist techniques. *Psychotherapy Research*, 5, 141–153.
- Hogue, A., Dauber, S., Liddle, H. A., & Samuolis, J. (2004). Linking session focus to treatment outcome in evidence-based treatments for adolescent substance abuse. *Psychotherapy: Theory, Research, Practice, Training*, 41, 83–96.
- Hogue, A., Liddle, H. A., & Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy*, 33, 332–345.
- Hogue, A., Liddle, H. A., Rowe, C., Turner, R. M., Dakof, G. A., & LaPann, K. (1998). Treatment adherence and differentiation in individual versus family therapy for adolescent substance abuse. *Journal of Counseling Psychology*, 45, 104–114.
- Hox, J. (2002). *Multilevel analysis techniques and applications*. Mahwah, NJ: Erlbaum.
- Hurlburt, M. S., Garland, A. F., Nguyen, K., & Brookman-Frazee, L. (2009). Child and family therapy process: Concordance of therapist and observational (this issue).
- Jensen, J. P., Bergin, A. E., & Greaves, D. W. (1990). The meaning of eclecticism: New survey and analysis of components. *Professional Psychology Research and Practice*, 21, 124–130.
- Jensen-Doss, A., Cusack, K. J., & de Arellano, M. A. (2008). Workshop-based training in trauma-focused CBT: An in-depth analysis of impact on provider practices. *Community Mental Health Journal*, 44, 227–244.
- Kataoka, S., Fuentes, S., O'Donoghue, V. P., Castillo-Campos, P., Bonilla, A., Halsey, K., et al. (2006). A community participatory research partnership: The development of a faith-based intervention for children exposed to violence. *Ethnicity and Disease*, 16, 89–97.
- Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 44, 1116–1129.
- Kazdin, A. E., Siegel, T. C., & Bass, D. (1990). Drawing on clinical practice to inform research on child and adolescent psychotherapy: Survey of practitioners. *Professional Psychology: Research and Practice*, 21, 189–198.
- Kazdin, A. E., & Weisz, J. R. (1998). Identifying and developing empirically supported child and adolescent treatments. *Journal of Consulting and Clinical Psychology*, 66, 19–36.
- Kazdin, A. E., & Weisz, J. R. (2003). *Evidence-based psychotherapies for children and adolescents*. New York, NY: Guilford Press.
- Knudsen, H. K., Johnson, J. A., & Roman, P. M. (2003). Retaining counseling staff at substance abuse treatment centers: Effects of management practices. *Journal of Substance Abuse Treatment*, 24, 129–135.
- Lambert, M., & Hill, C. (1994). *Assessing psychotherapy outcomes and processes. Handbook of psychotherapy and behavior change* (4th ed., pp. 72–113). New York, NY: Wiley.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Mayo, E. (1933). *The human problems of industrial civilization*. New York, NY: MacMillan.
- McLeod, B. D. (2001). *The therapy process observational coding system for child psychotherapy*. Los Angeles: University of California. Unpublished instrument.
- McLeod, B. D., & Weisz, J. R. (2008). The Therapy Process Observational Coding System for child psychotherapy strategies: Measure characteristics and prediction of outcome in usual clinical practice. (under review).
- Mufson, L. H., Pollack-Dorta, K., Olfson, M., Weissman, M. M., & Hoagwood, K. (2004). Effectiveness research: Transporting Interpersonal Psychotherapy for Depressed Adolescents (IPT-A) from the lab to school-based health clinics. *Clinical Child and Family Psychology Review*, 7, 251–261.
- National Advisory Mental Health Council. (1998). *Bridging science and service: A report by the National Advisory Mental Health Council's Clinical treatment and Services Workgroup*. Bethesda,

- MD: National Institutes of Health/National Institute of Mental Health.
- Nelson, T. D., Steele, R. G., & Mize, J. A. (2006). Practitioner attitudes toward evidence-based practice: Themes and challenges. *Administration and Policy In Mental Health*, 33, 398–409.
- Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting Clinical Psychology*, 75, 829–841.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100, 322–331.
- Schley, C., Ryall, C., Crothers, L., Radovini, S., Fletcher, K., Marriage, K., et al. (2008). Early intervention in the real world: Early intervention with difficult to engage, 'high-risk' youth: Evaluating an intensive outreach approach in youth mental health. *Early Intervention in Psychiatry*, 2, 195–200.
- Schoenwald, S. K., Heiblum, N., Saldana, L., & Henggeler, S. W. (2008). The international implementation of multisystemic therapy. *Evaluation & The Health Professions*, 31, 211–225.
- Sheehan, A. K., Walrath-Greene, C., Fisher, S., Crossbear, S., & Walker, J. (2007). Evidence-based practice knowledge, use, and factors that influence decisions: Results from an evidence-based practice survey of providers in American Indian/Alaska Native communities. *American Indian and Alaska Native Mental Health Research*, 14, 29–48.
- Vinnars, B., Barber, J. B., Norén, K., Gallop, R., & Weinryb, R. M. (2005). Manualized Supportive-Expressive Psychotherapy versus nonmanualized Community-Delivered Psychodynamic Therapy for patients with personality disorders: Bridging efficacy and effectiveness. *American Journal of Psychiatry*, 162, 1933–1940.
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, 61, 620–630.
- Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes". *Psychological Bulletin*, 122, 203–215.
- Weersing, V. R., Weisz, J. R., & Donenberg, G. R. (2002). Development of the therapy procedures checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child and Adolescent Psychology*, 31, 168–180.
- Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist*, 61, 671–689.
- Wells, K. B., Sturm, R., Sherbourne, C. D., & Meredith, L. S. (1996). *Caring for depression*. Cambridge, MA: Harvard University Press.
- West, S. G., Duan, H., Pequegnat, W., Gaist, P., Des Jarlais, D. C., Holtgrave, D., et al. (2008). Alternatives to the randomized control trial. *American Journal of Public Health*, 98, 1359–1366.
- Westfall, J. M., Mold, J., & Fagnan, L. (2007). Practice-based research—"Blue Highways" on the NIH roadmap. *The Journal of the American Medical Association*, 297, 403–406.
- Zima, B. T., Hurlburt, M. S., Knapp, P., Ladd, H., Tang, L., Duan, N., et al. (2005). Quality of publicly-funded outpatient specialty mental health care for common childhood psychiatric disorders in California. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 130–144.
- Zoffness, R., Garland, A., Brookman-Frazee, L., & Roesch, S. (2009). Case management as a significant component of usual care psychotherapy for youth with disruptive behavior problems. *Child & Youth Care Forum*, 38, 185–200.